Big data and the morality of inclusion

Michele Loi and Markus Christen

Introduction

In this chapter, we analyze questions pertaining to the use of predictive models based on big data from the point of view of Allen Buchanan's 'morality of inclusion" (Buchanan 1996; Buchanan, Cole, and Keohane 2011). Here we discuss to potential ethical risks of big data that are normally not discussed in the literature. The prevailing themes in the big data ethics literature are privacy, discrimination, and threats to autonomy, in particular autonomy as informational self-determination (Floridi 2012; Barocas and Nissenbaum 2014; Mittelstadt and Floridi 2015; Barocas and Selbst 2016; Christen et al. 2016; Vayena and Blasimme 2017). Here we show that beyond these concerns, big data also raise distinct issues that can be properly analyzed by using Buchanan's conceptual framework of the *morality of inclusion.* We illustrate ethical threats of big data for the morality of inclusion with two example of big data applications, one taken from the insurance domain and the other from the employment domain.

1. The morality of inclusion

One way to introduce the concept of the morality of inclusion is by discussing the idea of the social construction of disability (Oliver 1990; Amundson and Tresky 2008). Buchanan recognizes that disability results from the interaction of natural components (impairments in biological function) and social components (the context in which the function in question results inadequate). Thus,

whether any given trait *T* counts as a disability in a society S depends on features of that society: "it is only those inabilities that interfere with the performance of socially significant tasks (in one's social environment, relative to one's reference group) that count as disabilities" (Buchanan 1996, 38). For example, dyslexia, a biological dysfunction, is not a disability in a society that does not rely on written communication. Advanced numerical and verbal skills, such as they are necessary in technically sophisticated societies, imply the exclusion of mildly cognitively impaired individuals from many forms of interaction. Notice that not all disabilities fall in these categories. We do not refer to severe disabilities that involve extreme impairments in all main areas of development (psychomotor, emotional, communicative, social, and cognitive) and in a broad range of imaginable social contexts. Here, sophisticated medical and psychosocial support is a precondition to allow even for a basic level of social participation. But it has been argued that this framework can explain why even serious physiological impairments like blindness and deafness are to a significant extent disabling only by virtue of the way the social environment is constructed (Oliver 1990; Amundson and Tresky 2008; for a critique see Barclay 2011).

More generally, adopting specific institutions and technologies (and institutions *affecting* the adoption of specific technology) determines *the kind of persons who can contribute to society by cooperating with others and who cannot*. One example of this is provided by the largely theoretical discussion about the possibility of enhancing individuals beyond normal function through genetic engineering.[1] From Buchanan and others' points of view, normally functioning citizens (relative to current standards) living in an imagined future society in which most people will be genetically enhanced (compared to current levels) will be the 'new disabled'. This will happen if

[1] This domain of technology that has seen breakthrough advances in recent time with the CRISPR-CAS 9 technique (Harris 1998, 2007; Buchanan et al. 2000; Ledford 2015). More recently there have been news in the press concerning the fact that the first genetically enhanced humans (individuals with CRISPR-engineered AIDS resistance) may have allegedly already been created, but no official confirmation of this alleged breakthrough could be produced (Ryan n.d.).

citizens who are merely 'normal' (by current standards) will be excluded from forms of cooperation requiring capacities that count as 'enhanced', relative to the current normality standard.[2]

Buchanan uses the expression "dominant cooperative scheme" (Buchanan 1996, 40) to refer to the assemblage of institutional and technical factors that emerge from the cumulative interaction of several individuals over time and that determine who is more likely to be included in social cooperation. Our choices leading to one or a possibly different dominant cooperative scheme matter morally because:

> "Each individual also has an important interest in having access to a dominant cooperative scheme that is the most productive and rewarding form of interaction in which he or she can participate effectively" (Buchanan 1996, 42).

Buchanan denies that justice requires in every single instance the adoption of the most inclusive feasible dominant cooperative scheme. There are trade-offs between the features making a cooperative scheme more inclusive and other features resulting in greater human well-being. Thus, making a cooperative scheme inclusive to the extreme may be too costly for society. For example, our current cooperative scheme could more easily include the mildly cognitively disabled, by relying on more primitive productive arrangements (e.g., those not requiring advanced numerical and verbal skills). A valid moral argument against such cooperative scheme however is that it would plausibly not be able to produce the amount of resources that are needed to meet many important citizens' needs (including the need for education and health care also for mildly cognitively disabled persons). However, one can find a balance between reasons to make social cooperation more

[2] This view is entailed by, but does not entail, Daniels' that if normal functioning is defined as statistically normal functioning relative to one's contemporaries, individuals that are normal by current standard who qualify as having health needs. For objections against this view see (Loi 2013).

inclusive and reasons to promote other desirable features of social cooperation, which may pull in opposite directions. This requires *partially* adjusting features of social cooperation, at least locally, and to a given degree, assuming reasonable trade-offs.

Making reasonable adjustments to the dominant (institutional) scheme of cooperation even though it has some costs is morally analogous to adjusting the *physical* infrastructure of cooperation that decent societies consider ethically mandatory. This moral requirement appears to be widely socially supported. Our society already accepts the principle that some adjustments in the physical infrastructure of cooperation are ethically mandatory in that reasonable. For example, most people agree that society ought to provide for the special needs of disabled persons by building wheelchair ramps and that companies should make reasonable efforts to favor the inclusion of disabled individuals. These are necessary to ensure an otherwise unachievable inclusion of persons with disabilities in forms of social cooperation, deemed important for their flourishing as human individuals (Nussbaum 2006). There is also growing acceptance for the idea that companies or tax-payers (not disabled persons themselves) should subsidize the diffusion of technologies that enable workers with mild visual impairments to be better integrated in the work place (e.g. technologies to facilitate reading from computers) (Radu n.d.).

According to the morality of inclusion, unequal access to technological innovations may raise morality of inclusion problems, similar to the exclusion of the mildly disabled from social cooperation. The exclusion of the mildly disabled, just as the exclusion of the people lacking access to technology, is a distinct harm from severe deprivation and political domination (Buchanan, Cole, and Keohane 2011).

As we shall briefly show in what follows, exclusion from the dominant cooperative framework may also result from technologies that do not enhance traits above normality. Other technologies may contribute to raise the threshold of normal ability for cooperation, in a given

society. The analogy between choosing a more inclusive cooperation scheme and modifying the physical structure of cooperation is only partial, because society does not have the same amount of control on the ways in which people cooperate. Buchanan points out that although these are not entirely subject to human deliberate social choice, there might be moments "at certain critical junctures in human history" (Buchanan 1996, p. 41) where it may be able to exercise some degree of choice. Arguably, the diffusion of predictive models based on a large variety and volume of diverse data produced at considerable speed ("The Four V's of Big Data" n.d.) represents such a critical juncture. Thus, in what follows we consider the social implications of the diffusion of predictive models based on big data from the point of view of the morality of inclusion.

We try to explain why we consider the diffusion of predictive models based on big data a critical juncture that is relevant to the morality of inclusion, in the following two sections. To do so, we focus on applications of predictive models based on big data in two different domains: insurance and human resources.

2. Big data in insurance

Big data can be used in the insurance domain in many different ways. A recent report written by the Geneva Association (an international think tank bringing together experts from some of the world's wealthiest and most powerful insurance companies) states that:

> New sources of digital data, for example in online media and the Internet of Things, reveal information about behaviours, habits and lifestyles that allows us to assess individual risks much better than before. In many instances, better data

> makes it possible to better align premiums and risks and to reduce the overall cost of insurance. […] But arguably the greatest societal benefits come from the potential to reduce risks through better data and new digital technologies. […] By providing risk insights to policyholders, such 'digital monitoring' encourages behavioural change to reduce risks. Moreover, new data sources allow for the implementation of advanced risk management systems that use predictive analytics as a basis for early intervention and risk prevention. Ultimately, these technologies allow the role of insurance to evolve from pure risk protection towards predicting and preventing risks (Keller 2018, 6).

How does then a specific problem emerge from the point of view of the morality of inclusion (henceforth MoI) in the context of insurance? At a fairly abstract level, big data is problematic for MoI if the use of big data in insurance can negatively affect the ability of specific citizens to participate to social cooperation. This can happen if the following three conditions are satisfied:

1. There exist insurance products that are necessary for social inclusion.
2. The 'actuarially fair' premiums of those products are too high for some clients, making them uninsurable.
3. Society implicitly relies on private insurance companies to provide goods and services that are essential for social inclusion and does not foresee any social compensation mechanism for high costs clients.

We will present two hypothetical scenarios. In the first one, concerning physicians and surgeons, social exclusion seems morally objectionable also because of the poor accuracy of the predictions from which it ensues. In the second scenario, concerning car drivers, it is arguably morally objectionable in spite of being based (in the imagined case) on very accurate predictions.

Consider, then, the job of a surgeon or physician in a country with a legal obligation to purchase malpractice insurance. A big data approach for the analysis of physicians' risk could include, to begin with, many different sources of data directly relevant to the risk insured against, produced by surgeons' and physicians' in the course of their profession, e.g. the history of past interventions and diagnostic decisions that turned out to be correct or less so. It could also include many other sources of information. Let us imagine, for argument's sake, a world in which insurers access – with the consent of their clients – data produced by surgeons' cars (e.g. concerning their driving style), by social networks (e.g. to measure the frequency of swear words or harsh comments on one's friends' posts), from their credit cards and bank accounts (e.g. to measure how punctually a person pays her bills), academic scores for each year of school and university, data collected from online education platforms (including response speed during test taking and rate of correct vs. incorrect answers), and, finally, measurements by sensors of the Internet of Thing placed in hospitals, in the places where surgeons and physicians routinely move. All these data could theoretically contribute to a more accurate assessment of the behavioral and psychological traits of physicians, e.g. reliance on intuition vs. reflection, swiftness of judgment, accuracy, care for details, sensitivity to the needs of others, etc.. All this could contribute to a more accurate assessment of the risk insured against. Suppose that insurance against malpractice is compulsory and companies deny insurance to all physicians with a risk score higher than a certain threshold. Or perhaps, above such threshold, they charge a premium that is out reach for young physicians (except perhaps those who have families to support them). A sizeable proportion of surgeons or physicians that obtains an

unfavorable risk assessment (before ever exercising their profession) may be excluded from the profession.

Such practice may even turn out to be socially acceptable, because it would be perceived to promote patient safety, even though the statistical models used will be imperfect, like all statistical models are in reality. For example, suppose that insurance becomes quite high, and thus unaffordable for the group with a risk score above 0.7. (The meaning of a 0.7 risk score is, let us suppose, that 70% of the physicians with features found in that group are sued at least once in their career for malpractice.) Physicians or surgeons with high, or at least stable, income may be able to afford high premiums, as well as younger physicians from wealthy families that are willing to support them. Yet it may be possible that a group of young graduates fresh from medical school will be unable to bear such expense. It is also worth noting that 3 physicians every 10 who are singled out as high risk will be false-positives, i.e. people who in fact will never be sued for malpractice. Social exclusion concerns them too.

Notice that we provide this argument without any proof that the market equilibrium following from the increased availability of information will have exactly such effects on insurance prices. We only aim here to highlight that, as far as we know, this is a possibility that cannot be excluded on purely *a-priori* grounds. The point again is not whether physicians deserve or not to be excluded from their job. It is to bring attention to the fact that some individuals would be excluded – made disabled – by their data, even before they are given a chance to prove and exercise their skills.

Our second example concerns social inclusion, which is related to the transportation of workers. Suppose that liability insurance, more specifically the insurance against road liability that is obligatory in most countries for all drivers, will be based on sophisticated risk assessment models based on big data. Such insurance products, let us suppose, can predict the risk level of prospective

drivers based on data produced by, for example, social media (Aki 2018), email use, credit card purchases, etc., with high accuracy. Let us suppose for argument's sake that these data are demonstrably relevant to assessing risk, which provides a legal basis for collecting them. Now suppose that some prospective drivers are identified as the potentially very high risk (e.g. the worse centile in a normal risk distribution) and their insurance prices will be significantly above average. Not only they will be above the average for the insured population, but also significantly above the average premium (already quite high) nowadays paid by drivers who are young, inexperienced, (and male, in countries where gender discrimination in insurance is allowed), purchasing insurance for the first time. Suppose that we are dealing with a society without adequate public transportation means: some people *need* to use their car in order to get to work. These people would be excluded from their job – they would become the new disabled.

It may be objected that very distracted or impulsive drivers are already penalized given the existing arrangement in which prices are the highest for large group of clients, for example, young and inexperienced drivers. In response: in the current practice, drivers start with the highest premiums in the absence of driving history but it is difficult to identify the extreme end of the risk distribution. But temperamentally bad drivers have at least a chance to learn to drive carefully and improve their risk scores, compared to baseline predictions. In the big data scenario imagined here, big data allow to identify very high-risk drivers before they accumulate a substantial driving history. High-risk drivers may end up paying a significantly higher price compared to that young and inexperienced (male) drivers pay today on average.

In a way, this is simply consequence of distributing the burden of risk coverage more accurately, assuming constant costs for insuring the entire pool. The fact that some high-risk clients will pay higher premiums than they do today, even compared to the high premiums paid by young male inexperienced drivers, is simply the other side of the coin of the advantages of using big data

that are often presented as the reason to introduce these methods. It is often alleged that predictions based on big data have the potential to identify *low risk* clients that are incorrectly priced as *high risk* because they are young and inexperienced (and males, in jurisdictions in which gender information can be used for insurance purposes) (Marabelli et al. 2017). If that is true, assuming that the introduction of the new models does not affect the overall cost of insuring the pool,[3] it follows that the price reduction for young inexperienced drivers who are actually low risk, and will be paying less, will be compensated by a rise in the premiums of young inexperienced drivers who are actually high risk. In other words, the current, highly imperfect method for measuring the risk profile of young and inexperienced drivers is equivalent to implicit cross-subsidy from the actual low-risk to the benefit of the actual high-risk. The betterment of prediction methods allows a pricing that more accurately reflects the individual risk of individuals who are, now, considered uniformly high risk (e.g. based solely on age and driving history). But the more actuarially accurate pricing scheme may lead insurance premiums to levels that many high-risk drivers are unable to pay, if unsubsidized by others.

We argue that this is not only a concern of individual responsibility but also one of social morality. Or at least it is so when driving insurance is essential to reaching the workplace, and being singled out as a high-risk customer for liability insurance excludes individuals who need to be able to drive, from a wide range of possible jobs. The people predicted to be very high risk at the wheel may become the new disabled in such society.

In conclusion, the risk is that through predictions based on big data we end up being a society in which certain categories of people are excluded from cooperation – they become the new disabled. The morality of inclusion entails that we have *pro tanto* ethical reasons to avoid a society

[3] That may not be the case if feedback systems based on big data also reduce the overall cost of insuring the pool. If the cost-reduction is significant, it may be possible to lower premiums for the low risk without raising premiums for the high risk. We consider here a scenario in which the overall cost of insuring the pool does not change drastically.

such as this – where the reasons being *pro tanto* stands for their being reasons that need to be balanced against reasons (e.g. of efficiency, or safety) to arrive at a reasonable compromise.

3. Talent analytics

Let us now consider the second case study, that of talent analytics. By talent (or people) analytics one means the use of measurements and data, as opposed to common sense, general scientific considerations disconnected from actual data about individuals, and gut instinct, to ensure productivity, engagement and retention of high talent (Davenport, Harris, and Shapiro 2010, 2). Examples of people analytics are (Davenport, Harris, and Shapiro 2010):

- calculating the optimal number of staff members to deal with customers on the front desk;
- choosing the right kind of personality profile to deal with customers on the front desk;
- assessing the impact of health and wellness industry programs;
- measuring the ability to take initiative and using it to predict performance;
- using data when deciding personalized health and fitness programs for athletes, or for contract decisions. In the latter case, sport teams may try to predict risks inherent in hiring a promising athlete, who may become inactive due to accidents and other diseases;
- analyzing the flow of information between team members to improve communication and problem solving by groups;
- analyzing employee satisfaction through surveys;
- collecting and analyzing key performance data, to assess personal achievements and alignment with the company's objectives;

- analyzing turnover and business opportunity to predict shortages or excesses of human capabilities before they happen;
- developing indicators that predict how likely employees are to stay with the company, by identifying what employees value the most;
- optimizing a store's next day work schedule, based on predicted individual sales performance combined with other supply chain decisions.

People analytics is used by many large companies. For example, Harrah's (a hotel and casino chain) people analytics "observe and record the frequency with which customer-facing staff member smile, because that behavior is highly correlated with customer satisfaction" (Davenport, Harris, and Shapiro 2010, 5). Supporters of this practice argue forcefully that managers should be persuaded to adopt analytical decision making (Davenport, Harris, and Shapiro 2010). For example:

> In late 2009 Harrah's began recruiting an external sales force and used organizational psychologists to create a predictive assessment for the job. But during the interview process managers became emotionally attached to some of the candidates with low probabilities of success. The analysts were prepared: They used randomized testing to prove that analytics was the superior method, and relied on their interpersonal skills to sway decisions when necessary (Davenport, Harris, and Shapiro 2010, 6).

Here again notice that the goal of talent analytics appears fully legitimate in a business oriented and free market society. Who, as a customer, does not wish to be dealt by a more cheerful and polite front-desk employee? Who doesn't wish to do business with companies offering competitive prices

and yet constantly deploying a sufficient workforce to adequately respond to customer demand at all time? The idea that the most competent employees should be more likely to be hired and retained, especially at times of economic downturns when some layoffs are unavoidable, resonates with widespread ideas of meritocracy.

However, a significant downside of people analytics is that it may create a new form of social exclusion. This prospect becomes more problematic the greater the amount of data available about individuals and the more accurate and personalized predictions become. Imagine a world in which education produces many more data than it does today, e.g. because massive open online courses (MOOC) become more accepted as educational credentials, or because pedagogic innovation, such as the inverted classroom, include interactions with online teaching tools, producing huge amount of behavioral data about all pupils. Moreover, data from education can be aggregated together with data from other sources, such as social media, financial behavior (e.g. how punctually one pays one's bills) etc. Let us consider the objection that this would encounter a strong opposition from the data subject, which could be exercised through the tools afforded to individuals by the existing privacy legislation. But the right to privacy – understood the right to control information about oneself – could be formally respected because data may be collected with the consent of the data subjects, in a setting in which prospective employees have full control over the information they share and are fully aware of their predictive purposes. It is not entirely implausible to imagine that many people would consent to information from various sourced being used in a people analytics setting in a competitive setting. Some people may think that their personality, as reflected in the data, confers them a competitive advantage in the job market. Cognitive biases and a positive self-image may lead individuals to imagine a positive outcome of such data driven evaluations, as opposed to appropriately weighting the risk of being rated negatively. We know from experimental evidence

that people have a misguided, excessively optimistic perception of where they stand in the distribution of positive traits (Alicke et al. 1995; Kruger and Dunning 1999).

Imagine a scenario in which employers can select the employees that are statistically more likely to be extrovert, agreeable, open, conscientious and confident,[4] with specific emphasis on specific traits depending on the features of the job in question. They would also be able to exclude those whom they legitimately expect to be less productive than competitors or to be more problematic in ways that are relevant to organizations (e.g. in terms of their social skills and acceptance of workplace diversity). Suppose, now, that the predictive models developed by employers are able to identify the least suitable employees with enough accuracy during probation time or even during the selection process, e.g. as they would have access to the social network profiles of the candidates short listed for a role. This would minimize the expected costs for the company of making the wrong hire.

It is far from obvious that such practices would face strong social opposition. Anyway, this is a philosophical essay where social opposition should not be taken at face value but questioned in terms of the validity of its moral premises. Considering the hiring practice in itself, that is, in isolation from its broader societal implications, it is understandable that employers should use all the tools at their disposal in order to screen the best candidates. Moreover, the demand by customers to receive better services aligns with social utility and the selection of suitable candidates for that purpose with meritocratic ideas.

One argument against such practices derives from the morality of inclusion. The risk once again is that we move from the existing society which is able to accommodate even low-efficiency and disagreeable persons into its dominant scheme of cooperation, at least temporarily. This occurs

[4] These correspond to the so-called 'big five' factors of personality psychology, with 'confidence' considered as the quality at the opposite end of the spectrum relative to the 'neuroticism' factor. Michal Kosinsky and others have shown that personality profiles based on the analysis of Facebook Likes have higher internal and external validity than those by humans (Youyou, Kosinski, and Stillwell 2015)

also thanks to a natural veil of ignorance that makes it difficult for employers to identify, and avoid, all the most problematic profiles. In other words, thanks to the current low predictive efficacy, the economic and social burden of interacting with low performers and cooperators with problematic psychological traits is shared between different organizations, that hire and discard such workers without the ability to collect the data necessary to identify and avoid them once and for all.

Even in societies with market friendly labor laws, where it is relatively simple to dismiss workers, and especially the worst workers are in fact dismissed more often than others, it may take time before a publicly accessible track record of some workers being less suitable than most others, showing that a given individual is a high-risk hire and not the best choice even for a humble and low salary job. But the most interesting difference brought about by big data may involve hiring itself. Suppose that societies with very strict labor laws, making dismissals extremely difficult legally, will not protect the high-risk workers from the use of big data analytics *before* they are hired as employees. In such context, the harder it is to dismiss an underperforming worker, the higher the degree of caution exercised in hiring – the degree of risk an employer will be willing to accept.

A possible consequence is that individuals with traits that are undesirable for a wide range of jobs will be *permanently* excluded from employment, especially in a context of generally high unemployment. Persons with personality traits such as a disposition to exert low effort, bad temperament, or prejudices may even be excluded from *every* job, including low-skill jobs, apprenticeships or early career jobs necessary to acquire a minimal level of skill necessary for any job. Of course, whether this scenario realizes or not, depends also on the overall level of unemployment, which in turn is affected by both macro-economic factors and the capacity of technology (in particular so-called AI) to cheaply replace certain skills of workers, increasing the demand and market value of other skills, that not all workers may currently possess (Autor 2015). It also depends on the distribution of the traits in question in the population: if some desired traits

are hard to find, employers will have to accept hiring individuals with accompanying undesired traits. Yet it seems hard to exclude *a-priori* the possibility that, at least with the high unemployment levels of many current economies, the people at one extreme of the distribution of undesired traits (according to the prediction based on their big data) may in fact be excluded from any sort of apprenticeships and jobs.

If some human qualities (e.g. of personality) become increasingly predictable through big data, individuals with poor probabilities of having qualities required for any job may find themselves cut out from all job opportunities from the beginning to the end of their potential working life. This will be possible if predictive models based on big data enable employers to make predictions about traits that are valued across a large spectrum of jobs. Those lacking those general-purpose traits would become, one could say, the new disabled.

4. Does the morality of inclusion single out a distinct moral concern?

The ethical downsides of big data have not been entirely ignored. For example, the above-mentioned report of the Geneva association about the implications of big data in the insurance domain reports mentions concerns of privacy, individualization of predictions and competition. Much has been written about the risk of *discrimination* resulting from big data, including, in the insurance domain, the issue higher premiums or reduced access to insurance products, particularly for traditionally discriminated groups (Zarsky 2014; Barocas and Selbst 2016). Notice however that sorting individuals into risk classes for the sake of adjusting premiums and excluding high-risk clients from services is a common and apparently accepted practice in the private insurance context. Insurance premiums determined according to adequately measured relevant risks are often described as 'fair' (more precisely, actuarially fair). The principles of actuarial fairness find widespread acceptance in societies where markets play a substantial role in allocating resources, as

shown by current laws and policies concerning such markets, at least for insurance that does not cover services connected with basic needs and human rights, that is to say for most insurance markets excluding health insurance affecting access to essential levels of health care (Meyer 2004, 31). Certainly, some *indirect discrimination* effects, for example zoning (also called redlining) are considered controversial in some societies. For example, selling more expensive insurance products to residents in neighborhoods where there is a higher prevalence of persons from ethnic minorities is widely considered unjustly discriminatory in the United States (Barocas and Selbst 2015; Zarsky 2014).

Notice, however, that not all forms of discrimination based on big data may be strongly be correlated with ethnic minority status, gender or race. Moreover, as legal scholar Tal Zarsky points out in analyzing indirect discrimination, it is unclear that many of the ethical reasons that make indirect discrimination problematic apply to the insurance case across all domains (Zarsky 2014, 1400). The traditional problem of indirect discrimination, moreover, concerns *socially salient* groups (Lippert-Rasmussen 2007), where by definition a group is socially salient "if perceived membership of it is important to the structure of social interactions across a wide range of social contexts" (Lippert-Rasmussen 2007, 386). Not all discriminating profiles based on big data will be associated with membership to a group that is socially salient, such as race, gender, ethnicity or age. Still, some big data methods that are not associated with socially salient groups may identify high-risk individuals and exclude them from cooperation, in ways that are permanent or almost so.

The real novelty of predictions based on big data could be the increase of finer-grain discrimination between different sub-types of members of the socially salient groups that have been traditionally discriminated against. For example, in civil liability insurance for road accidents it may be possible to better distinguish males who are actually expensive to insure vs. males who are actually inexpensive to insure. More fine-grained and personalized statistical proxies of risk, based

on behavioral data, may even replace traditional social categories (e.g. young and old) still used in order to discriminate (fairly, in the actuarial meaning of 'fair'). Similar considerations may be applicable to the domain of employment, where the use of statistical predictions may be perceived as providing a fairer way to allocate the chance of being unemployed, compared to current practices.

Since it is not obvious that predictive algorithms will be discriminatory in the current sense, it is important to stress the concern of the morality of inclusion differs from moral objections that derive from the principle of non-discrimination. It also differs with the moral concern fairness in a broader sense, often defined as the concern that individuals should not be penalized due to factors for which they cannot be considered responsible (Avraham, Logue, and Schwarcz 2014; Dworkin 1981; Cohen 1989). Some of the statistical proxies that may lead to exclusion from social cooperation – due to their impact on insurance or employment risk assessment – may be behavioral profiles based on behaviors that high-risk individuals can control and for which they can be considered responsible.

5. Political morality and policy implications

The morality of inclusion is not (yet?) widely accepted as a source of normative requirements, unlike more publicly recognized aspects of social morality, such as the moral duty to meet urgent needs (Scanlon 1975), the requirement to avoid discrimination and be fair. The paradigmatic case Allen Buchanan uses to illustrate the attraction of a morality of inclusion concerns mild physical and mental disabilities. Unlike the disability case, the people at risk of social cooperation by virtue of predictions that can be made about them may often be categorized as lacking those qualities. This may include moral virtues, which most citizens find reasonable to *demand* of workers with whom

they interact. Supporters of meritocracy may even welcome big data approaches that distribute chances of unemployment in a less random fashion compared to current affected by arbitrary intuitive human assessments. It is also not so clear how strong solidarity can be felt towards categories of people who do not appear to be disabled, mildly or severely, in the medically established sense. These people may be considered to be, on account of their 'flaws' less 'deserving' of obtaining the 'privilege' to have a job, especially in high unemployment contexts in which the very fact of obtaining a job may be considered a privilege.

Thus, the moral idea of a morality of inclusion single out a specific aspect making the adoption of such technologies problematic, independent from their impact on fairness, discrimination, or meeting basic needs. Arguments that rely on a duty to include would stress the responsibility of persons responsible for development and adoption of such technologies to avoid a historical juncture in which certain minorities of people are identified as carriers of undesirable traits and, on account of those traits, be excluded from cooperation altogether. This is a distinct type of interest, compared to extreme deprivation and unfair discrimination. The problem of deprivation associated with unemployment can arguably be dealt with by providing an income sufficient to meet the most urgent basic needs, such as adequate nourishment, shelter, etc. The specific interest in social inclusion in productive cooperation, that is in question here, needs to be clearly singled as such, because many forms of welfare that may be designed to cope with the transformation led by big data and analytics could fail to address it, if it only targets severe deprivation. It is important to emphasize that the interest of a person which is in question is that in joining "the most productive and rewarding form of interaction in which he or she can participate effectively" (Buchanan 1996, 42) and this is arguably not even approximated by providing individuals with income unrelated to the performance of a job and to the reasonable expectation of obtaining one in the future.

As Allen Buchanan acknowledged (Buchanan 1996), there may be situations in which it is not *reasonable* to meet each citizen's interest of inclusion fully. Nowadays, where many simple tasks are performed by machines and many complex tasks require sophisticated cognitive capacities, it is arguably more difficult for the mildly cognitively disabled to find a role *qua* cooperators, yet few would endorse a transition to a society in which cognitive skills were less important than they are today, at the cost of a significant worsening the average standard of living. Still, as Buchanan observes, the interest of inclusion is legitimate at least *pro tanto* and decent societies should try to balance it with other reasonable moral demands.

Notice, however, that the condition of the new disabled in the big data world - the people unable to work because of unfavorable predictions made about them - may be even worse than that of people who are disabled in the traditional sense, e.g. due to biological physical or mental impairments. For unlike the latter, the subjects of unfavorable predictions are not considered physically or mentally impaired. Thus, they may not elicit solidarity from would be cooperators. Maybe for efficient and socially adequate workers it is easier to sympathize with persons with recognized disabilities, than with apparently healthy workers lacking some important socially desirable traits. Moreover, the lack of certain psychological and moral qualities in employees, e.g. kindness, sociability, or openness towards diversity, may negatively affect the work environment generating hatred towards the workers lacking those qualities. This is one more reason to expect that measures to promote the inclusion of the 'new disabled' are not likely to receive widespread support in society.

Yet we believe that a persuasive case could be made for considering this outcome as morally unsatisfactory. The challenge is to identify a policy that could achieve this in a socially acceptable way. Let us consider people analytics, first. The most traditional policy for protecting workers, namely a prohibition to lay off people, may actually backfire, in so far as it will raise the value for

employers to have predictions and reduce even further their propensity to hire a person without favorable predictions. Another possible policy would be to 'artificially' boost the rate of economic growth, e.g. through state expenditure or tax cuts, so as to stimulate employment. The expectation would be that as the unemployment rate goes down, employers cannot afford to be too choosy and will eventually hire also individuals with unfavorable predictions. One difficulty with such solution is that boosting growth with macro-economic policy comes at a cost, e.g. society may have to spend more to pay interests on debt and this, especially for debt-ridden states, may be a considerable cost that society reasonably wants to avoid. This leaves two policy options on the table: first, to prohibit the use of people analytics both in decisions concerning whom to lay off and whom to hire; second, the state as employer of last resort.

First, let us consider the prohibition. One may restrict the use of big data predictions to formative contexts (i.e., training of people) rather that selection contexts (i.e., employment decisions). This argument relies on three premises: First, many of the negative personality traits that may be unfavorable in certain work conditions are not static and deterministic. For example, people are able to improve social deficits or to develop coping strategies when those deficits become integrated in training programs. For example, in medicine, teaching medical students increasingly includes training of social competences such as empathic concern (Katsarov et al., in preparation). Second, there is good evidence to assume that predictors of complex personality traits will have limits in their explanatory power, even when they rely on big data. Decades of experiences in test psychology have shown the difficulty to make psychometric measurements ecologically valid; i.e. there is reason to believe that also big data predictions will be to a significant degree unreliable. In a formative context, it will be much easier to cope with those types of imprecision compared to using them for job selection. Third, at least some of the character traits that may be used for selection decisions are constitutive for the self-understanding of individuals and traits from which it

is not reasonable to expect individuals to dissociate themselves from. For example, limited empathic concern – although negative in various job settings – may be part of a personality that overall performs well. The film character "Dr. House" may serve as an example here – an extremely asocial physician (who for sure would be sorted out by any "big data sociality detector") who nevertheless saved many patients. Thus, there is normative reason that speaks against the homogeneity that the selective use of big data predictions may generate. Nevertheless, formative use in training and continuous education allows to deal with imprecision and "special cases", making a restriction to this context plausible.

The second possibility would be the state as an employer of last resort. For example, local communities could use part of their funds (from taxpayer money) in order to hire people who in a people analytics context are rejected by all private companies. The policy comes at a cost because, like the former solution, it is certainly going to have an efficiency cost. If the goal of local communities is delivering services to the public, then it would be just as rational for public and private employers alike to avoid hiring individuals predicted to be less suitable for the relevant tasks, compared to workers selected using big data. However, one of the goals of such public organizations would be to promote the inclusion of these individuals. Relative to *this goal*, the policy in question will not be irrational.

A third possibility could be a combination of the former two. One may further imagine employment by the state that, beside inclusion, serves educational and research purposes as well. Educational purposes may include ways for enhancing the missing skills. It may still be possible to address the shortcomings of persons with unfavorable predictions combining work and training within employment by the state. Research purposes may be included also in order to produce data about what features are more malleable can be effectively modified, as a way of redressing non-employability determined by unfavorable predictions. Society has a long-term interest in

understanding which environmental measures can offset or at least mitigate the problem. The long-term interest in question is, again, not just economic efficiency but being a morally decent society that fulfills the obligations of the morality of inclusion.

Let us now turn to the insurance case. This appears to be slightly more manageable than the one concerning employment. Consider for example the case of compulsory insurance for medical malpractice. Here society may respond to the problem, by abolishing compulsory insurance against malpractice. Instead, the victims of medical malpractice could be compensated through a fund subsidized by taxpayers (or by the users of medical services, to achieve some kind of proportionality). For transportation, the insurance costs of drivers who are uninsurable by free markets could be subsidized by taxpayers. Eligibility for such subsidy may also be combined with a requirement to go through additional training, to acquire risk-minimizing strategies. Or public services could be made more efficient and capillary, making driving a personal car less of a necessity for social integration.

Conclusion

In this paper we have identified and described a dimension of the moral assessment of big data technologies. We have argued that it is conceptually independent from the moral obligation to avoid discrimination, unfairness, and severe deprivation. This concern is the morality of inclusion, that reflects the interest, of every individual, of participating in "the most productive and rewarding form of interaction in which he or she can participate effectively" " (Buchanan 1996, 42). From the point of view of the morality of inclusion, big data may be problematic because it enables more accurate predictions about individuals. This then can lead to their exclusion from cooperation.

We have indicated two mechanisms through which this may happen: first, individuals may be excluded from insurance products that are necessary to gain access to certain job opportunities; second, individuals may be excluded from all or most jobs by virtue of predictions made about their future productivity and their social attitudes and skills. We have explained the nature of the problem, illustrated the differences with discrimination and fairness issues, and argued that protection for the interest in question may be an unpopular political cause.

Since the morality of inclusion stands for an authentic interest of individuals, societies should take reasonable measures to mitigate the problem of big-data driven exclusion. In the final section of the paper we discuss four possible political strategies for people analytics: strengthening labor law in the direction of making workers dismissal more difficult, reduction of unemployment through macro-economic policy, prohibition of the use of people analytics in hiring and dismissing decisions, and the state as employer of last resort, which combines employment with training, education and research. We argue that the first can be counterproductive, the second may not be economically sustainable in many cases, and this gives societies reason to seriously consider the other two. For big data predictions in the insurance sectors there seems to be three equally reasonable strategies, namely, the abolition of compulsory insurance that causes the exclusion of some workers (where the state takes charge for compensating the victims), subsidizing policies for high-risk individuals, and adopting structural changes that reduce the dependency of workers on private insurance products that are essential for social inclusion.

References

Aas, Sean, and David Wasserman. 2015. “Natural and Social Inequality: Disability and Fair Equality of Opportunity.” *Journal of Moral Philosophy*, October. https://doi.org/10.1163/17455243-46810042.

Aki, Jimmy. 2018. “Is Your Insurance Company Snooping on Your Social Media Profiles?” *Money Check* (blog). September 25, 2018. https://moneycheck.com/insurers-social-media/.

Alicke, Mark D., Mary L. Klotz, David L. Breitenbecher, Tricia J. Yurak, and Debbie S. Vredenburg. 1995. “Personal Contact, Individuation, and the Better-than-Average Effect.” *Journal of Personality and Social Psychology* 68 (5): 804.
Amundson, Ron, and Shari Tresky. 2008. “Bioethics and Disability Rights: Conflicting Values and Perspectives.” *Journal of Bioethical Inquiry* 5 (2): 111–23. https://doi.org/10.1007/s11673-008-9096-3.
Autor, David H. 2015. “Why Are There Still So Many Jobs? The History and Future of Workplace Automation.” *The Journal of Economic Perspectives* 29 (3): 3–30.
Avraham, Ronen, Kyle Logue, and Daniel Schwarcz. 2014. “Understanding Insurance Antidiscrimination Laws.” *S. Cal. L. Rev.*, January. https://scholarship.law.umn.edu/faculty_articles/576.
Barclay, Linda. 2011. “Justice and Disability: What Kind of Theorizing Is Needed?” *Journal of Social Philosophy* 42 (3): 273–87. https://doi.org/10.1111/j.1467-9833.2011.01533.x.
Barocas, Solon, and Helen Nissenbaum. 2014. “Big Data’s End Run around Anonymity and Consent.” In *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, edited by Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum. New York, NY: Cambridge University Press.
Barocas, Solon, and Andrew D. Selbst. 2015. “Big Data’s Disparate Impact.” SSRN Scholarly Paper ID 2477899. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2477899.
———. 2016. “Big Data’s Disparate Impact.” *Cal. L. Rev.* 104: 671.
Buchanan, Allen E. 1996. “Choosing Who Will Be Disabled: Genetic Intervention and the Morality of Inclusion.” *Social Philosophy and Policy* 13 (2): 18–46. https://doi.org/10.1017/S0265052500003447.
Buchanan, Allen E., Dan W. Brock, Norman Daniels, and Daniel Wikler. 2000. *From Chance to Choice*. Cambridge, UK: Cambridge University Press.
Buchanan, Allen E., Tony Cole, and Robert O. Keohane. 2011. “Justice in the Diffusion of Innovation.” *Journal of Political Philosophy* 19 (3): 306–32. https://doi.org/10.1111/j.1467-9760.2009.00348.x.
Christen, Markus, Josep Domingo-Ferrer, Bogdan Draganski, Tade Spranger, and Henrik Walter. 2016. “On the Compatibility of Big Data Driven Research and Informed Consent: The Example of the Human Brain Project.” In *The Ethics of Biomedical Big Data*, edited by Brent Daniel Mittelstadt and Luciano Floridi, 199–218. Law, Governance and Technology Series. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33525-4_9.
Cohen, G. A. 1989. “On the Currency of Egalitarian Justice.” *Ethics* 99 (4): 906–44.
Davenport, Thomas H., Jeanne Harris, and Jeremy Shapiro. 2010. “Competing on Talent Analytics.” *Harvard Business Review* 88 (10): 52–58.
Dworkin, Ronald. 1981. “What Is Equality? Part 2: Equality of Resources.” *Philosophy and Public Affairs* 10 (4): 283–345.
Floridi, Luciano. 2012. “Big Data and Their Epistemological Challenge.” *Philosophy & Technology* 25 (4): 435–37. https://doi.org/10.1007/s13347-012-0093-4.
Harris, John. 1998. *Clones, Genes and Immortality: Ethics and the Genetic Revolution*. Oxford University Press, Oxford.
———. 2007. *Enhancing Evolution: The Ethical Case for Making Better People*. Princeton, NJ: Princeton University Press.
Keller, Benno. 2018. “Big Data and Insurance: Implications for Innovation, Competition and Privacy.” The Geneva Association—International Association for the Study of Insurance Economics. https://www.genevaassociation.org/sites/default/files/research-topics-document-

type/pdf_public/big_data_and_insurance_-_implications_for_innovation_competition_and_privacy.pdf.
Kruger, Justin, and David Dunning. 1999. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments.” *Journal of Personality and Social Psychology* 77 (6): 1121–34. https://doi.org/10.1037/0022-3514.77.6.1121.
Ledford, Heidi. 2015. “CRISPR, the Disruptor.” *Nature* 522 (7554): 20–24. https://doi.org/10.1038/522020a.
Lippert-Rasmussen, Kasper. 2007. “Nothing Personal: On Statistical Discrimination*.” *Journal of Political Philosophy* 15 (4): 385–403. https://doi.org/10.1111/j.1467-9760.2007.00285.x.
Loi, Michele. 2013. “You Cannot Have Your Normal Functioning Cake and Eat It Too.” *Journal of Medical Ethics*, January. https://doi.org/10.1136/medethics-2012-100989.
Marabelli, Marco, Sean Hansen, Sue Newell, and Chiara Frigerio. 2017. “The Light and Dark Side of the Black Box: Sensor-Based Technology in the Automotive Industry.” *Communications of the Association for Information Systems* 40 (1). http://aisel.aisnet.org/cais/vol40/iss1/16.
Meyer, Roberta B. 2004. “The Insurer Perspective.” In *Genetics and Life Insurance: Medical Underwriting and Social Policy*, edited by Mark A. Rothstein, 28–47. Cambridge MA: MIT Press.
Mittelstadt, Brent Daniel, and Luciano Floridi. 2015. “The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts.” *Science and Engineering Ethics* 22 (2): 303–41. https://doi.org/10.1007/s11948-015-9652-2.
Nussbaum, Martha Craven. 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, Mass: The Belknap Press.
Oliver, Michael. 1990. *The Politics of Disablement : A Sociological Approach*. New York: St. Martin’s Press.
Radu, Sintia. n.d. “Who’s Paying for Assistive Technology?” US News & World Report. Accessed January 2, 2019. https://www.usnews.com/news/best-countries/articles/2017-12-01/assistive-technology-keeps-growing-but-paying-for-it-is-next-challenge.
Ryan, Jackson. n.d. “Scientists in China Claim to Have Created First Gene-Edited Human Babies.” CNET. Accessed January 2, 2019. https://www.cnet.com/news/scientists-in-china-claim-to-have-created-first-gene-edited-human-babies/.
Scanlon, T. M. 1975. “Preference and Urgency.” *The Journal of Philosophy* 72 (19): 655–69. https://doi.org/10.2307/2024630.
“The Four V’s of Big Data.” n.d. IBM Big Data & Analytics Hub. Accessed January 2, 2019. https://www.ibmbigdatahub.com/infographic/four-vs-big-data.
Vayena, Effy, and Alessandro Blasimme. 2017. “Biomedical Big Data: New Models of Control Over Access, Use and Governance.” *Journal of Bioethical Inquiry*, October, 1–13. https://doi.org/10.1007/s11673-017-9809-6.
Youyou, Wu, Michal Kosinski, and David Stillwell. 2015. “Computer-Based Personality Judgments Are More Accurate than Those Made by Humans.” *Proceedings of the National Academy of Sciences* 112 (4): 1036–40. https://doi.org/10.1073/pnas.1418680112.
Zarsky, Tal Z. 2014. “Understanding Discrimination in the Scored Society.” *Washington Law Review* 89: 1375.